REMARKS

Claims 78-85, 96-103, 114-121, 132-139, previously withdrawn are now cancelled. Claims 68, 86, 104 and 122 have been amended, as suggested by the Examiner, to recite that the variant "has alpha-amylase activity." Claims 68, 86, 104 and 122 have also been amended to remove the phrase "(using SEQ ID NO:3 for numbering)" as this is redundant in view of the claim language "at position equivalent to positions...in SEQ ID NO:3."

Claims 140-155 have been added. New Claims 140-147 correspond to the allowable subject matter identified on page 10 of the Office Action. In particular, on page 10 of the Office Action, the Office states that claims 76, 77, 94, 95, 112, 113, 130 and 131 were objected as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

New claims 148-155 are directed to variants a parent a *Bacillus stearothermophilus* alpha-amylase wherein the variant is at least 90% or at least 95% homologous to the parent *Bacillus stearothermophilus* alpha-amylase and comprises one of the claimed pair wise deletions. Support for these claims is found, inter alia, in the specification at page 10, third and fifth paragraphs, and page 13, lines 13-27. Claims 148-155 are also believed to be allowable for the reasons supporting the allowability of new claims 140-147.

The specification is amended, as requested by the Examiner, to identify the application no. for U.S. Patent No. 6,093,562, and to provide the status for USSN 09/354,191.

Certified copies of the four Danish priority documents are enclosed.

An information disclosure statement is submitted herewith.

It is respectfully submitted that the present amendment presents no new issues or new matter and places this case in condition for allowance. Reconsideration of the application in view of the above amendments and the following remarks is requested.

I. The Rejection of Claims 68-72, 86-90, 104-108 and 122-126 under 35 U.S.C. 112

Claims 68-72, 86-90, 104-108 and 122-126 are rejected under 35 U.S.C. 112, first paragraph, as lacking adequate written description support. The Office states that the specification does not adequately describe the claimed genus of variants of a parent alpha-amylase having at least 80% identity to the referenced alpha-amylase and having a deletion of amino acid residues equivalent to positions 180 and 181, 179 and 181, 179 and 182 or 180 or 182. The Office suggests amending claims 68, 86, 104 and 122 to recite "wherein said variant has alpha-amylase activity, has at least 80%...."

Applicants have amended the claims as suggested by the Examiner to recite that the variant has alpha-amylase activity, and respectfully submit that claims 68-72, 86-90, 104-108 and 122-126 now fulfill the written description requirement of the Patent Code.

The test for determining compliance with the written description requirement is whether the disclosure of the application reasonably conveys to the artisan that the inventor had possession of the claimed subject matter at the time the original application was filed. See *In re Kaslow*, 217 USPQ 1089, 1096 (Fed. Cir. 1983). A specification complies with the written description requirement if it provides "a precise definition, such as by structure, formula, chemical name, or physical properties of the claimed subject matter sufficient to distinguish it from other materials." See, e.g., *University of California v. Eli Lilly and Co.*, 43 U.S.P.Q.2d 1398, 1404 (Fed. Cir. 1997); *Enzo Biochem v. Gen-Probe Inc.*, 63 U.S.P.Q.2d 1609, 1613 (Fed. Cir. 2002). A description of a claimed genus may be achieved by recitation of a representative number of species falling within the scope of the genus or by a recitation of structural features common to the members of the genus which constitute a substantial portion of the genus. See *Eli Lilly and Co.*, 43 U.S.P.Q.2d at 1569.

As presently amended, claims 68-72, 86-90, 104-108 and 122-126 now clarify that the variants have alpha-amylase activity. It is respectfully submitted that the disclosure of the present application provides a sufficient written description such that one skilled in the art would conclude that the Applicants were in possession of the claimed alpha-amylase variants at the time of filing. As disclosed in the specification, the claimed deletions can be applied to alpha-amylases having "at least 80% homology" to the reference sequences. See, e.g., the specification at page 4, lines 7-14. Describing a polypeptide by its sequence similarity to a reference sequence is well-accepted in the art as a suitable method for describing the polypeptide's structure and function, as polypeptides having very high degree sequence similarity of at least 80% homology to a reference polypeptide, are reasonably expected to have both a very similar structure and function/activity as the reference polypeptide.

Applicants have also described that the deletions are applicable to other alpha-amylases besides the reference sequences. Such alpha-amylases include variant alpha-amylases that have been altered at other positions, as described in the specification at pages 12-17.

In sum, Applicants' description that the claimed invention is applicable to those polypeptides which have a degree of homology of at least 80% to the reference sequence and Applicants' identification of representative examples, is sufficient for the artisan to conclude that Applicants had possession of the claimed invention as it provides recitation of structural features

common to the members of the genus and representative members falling within the claimed genus.

For the foregoing reasons, Applicants submit that the claims overcome this rejection under 35 U.S.C. 112. Applicants respectfully request reconsideration and withdrawal of the rejection.

II. The Rejection of Claims 68-75, 86-93, 104-111 and 122-129 under 35 U.S.C. 112

Claims 68-75, 86-93, 104-111 and 122-129 are rejected under 35 U.S.C. 112, as lacking enablement. This rejection is respectfully traversed.

Section 112 of U.S. Patent Code requires that the specification be "enabling" to a person skilled in the art to which the invention pertains. Enablement requires that the specification teach those in the art to make and use the invention without undue experimentation. *In re Wands*, 8 USPQ2d 1400 (Fed. Cir. 1988).

It is respectfully submitted that the claims are enabled for alpha-amylase variants which are at least 80% homologous to the reference sequence and comprise one of the pair wise deletions recited in the claims. The specification provides examples of representative species falling within the scope of the claimed invention in the form of suitable alterations which can be made in combination with the claimed pair wise deletions. See the specification at pages 12-17. In addition to these representative examples, as of the filing date of the present application, many *Bacillus stearothermophilus* alpha-amylases variants were known in the art, which provided information on important conserved sequences as well as sequences which could be altered without destroying the alpha-amylase activity. See, e.g., Holm et al., "Random mutagenesis used to probe the structure and function of Bacillus stearothermophilus alpha amylase", Prot. Eng., Vol. 3, No. 3, pp. 181-191 (1990), Vihinen et al., "Site-directed mutagenesis of a thermostable alpha-amylase from Bacillus stearothermophilus: putative role of three conserved residues, J. Biocem, 107(2):267-72 (1990), attached as Exhibits A and B, respectively.

Consistent with the Holm et al. and Vihinen et al. publications, the specification discloses that alpha-amylases falling within the scope of the claimed invention may be prepared by site directed mutagenesis and random mutagenesis, including localized random mutagenesis. See the specification at page 18-24. Although the production and identification of suitable variants falling within the scope of the claimed 80% homology would involve some experimentation, it is respectfully submitted that such experimentation is not undue. In particular, both the task of producing such sequences and the task of screening the sequences for variants with alpha-amylase activity and/or improved alpha-amylase activity was routine in the art as of the time of the

invention. *See id.* The artisan could, e.g., rapidly produce and screen thousands of alpha-amylase variants and identify those variants which possess alpha-amylase activity using the processes of site directed mutagenesis, random mutagenesis and localized random mutagenesis following by an alpha-amylase activity assay as described in the specification at page 54 ("Determination of α-amylase activity").

In addition, although the three-dimensional models available at the time of the present invention were generally fungal based alpha-amylase models, which had some significant limitations in the application to *Bacillus stearothermophilus* alpha-amylases, they did provide some insight into conserved structural features common to alpha-amylases in general. See Holm et al. and Vihinen et al. (modeling a *Bacillus stearothermophilus* alpha-amylase on the basis of the fungal TAKA alpha-amylase structure). Thus, although these models were not accurately representative of the complete structure of *Bacillus stearothermophilus* alpha-amylases, they did provide some guidance into a number of conserved structural features of the molecule which an artisan would preserve when making other alterations.

Accordingly, it is respectfully submitted that the present invention is enabled for making alpha-amylase variants having a claimed pair wise deletion and which are at least 80% homologous to the reference sequence.

For the foregoing reasons, Applicants submit that the claims overcome this rejection under 35 U.S.C. 112. Applicants respectfully request reconsideration and withdrawal of the rejection.

III. Conclusion

In view of the above, it is respectfully submitted that all claims are in condition for allowance. Early action to that end is respectfully requested. The Examiner is hereby invited to contact the undersigned by telephone if there are any questions concerning this amendment or application.

Respectfully submitted,

Date: October 6, 2004

Jason I. Garbell, Reg. No. 44,116
Novozymes North America, Inc.
500 Fifth Avenue, Suite 1600
New York, NY 10110
(212)840-0097

Protein Engineering vol.3 no.3 pp.181–191, 1990

Random mutagenesis used to probe the structure and function of *Bacillus stearothermophilus* alpha-amylase

Liisa Holm, Anu K.Koivula, Päivi M.Lehtovaara, Ari Hemminki and Jonathan K.C.Knowles

Biotechnical Laboratory, VTT, Tietotie 2, SF-02150 Espoo, Finland

Mutations that cover the sequence of *Bacillus stearothermophilus* α-amylase were produced by an efficient *in vitro* enzymatic random mutagenesis method and the mutant α-amylases were expressed in *Escherichia coli*, which also secreted the product. Ninety-eight mutants were identified by sequencing and their enzyme activities were classified into three classes: wild-type, reduced or null. A molecular model of the enzyme was constructed using the coordinates of Taka-amylase A and a consensus alignment of mammalian, plant, and bacterial α-amylases. The location of mutant amino acids on the model indicate that mutations which destroy or decrease the catalytic activity are particularly clustered: (i) around the active site and along the substrate-binding groove and (ii) in the interface between the central α/β barrel and the C-terminal domain. Exposed loops are typically tolerant towards mutations.

Key words: homology modelling/sequence alignment/starch

Introduction

α-Amylases (α-1,4-glucan-4-glucanohydrolase, EC 3.2.1.1) are widely distributed starch-degrading enzymes found in plants, animals and bacteria. Because of the importance of starch as a raw material in a number of industries, particularly the food industries, α-amylase is an enzyme of considerable commercial significance. Despite this, there is relatively little information concerning either structure or function of this interesting family of enzymes. It is a common observation that even the availability of a good three-dimensional structure does not necessarily make it possible to predict the effects of specific amino acid changes, even in the case where information about the enzyme mechanism is available (Knowles, 1988). For this reason we have developed a novel random mutagenesis method (Lehtovaara *et al.*, 1988). The method permits the enzymatic generation of a library of single (and multiple) base mutations throughout a defined region of DNA which can be several kilobases long. This method has now been applied to the cloned gene of *Bacillus stearothermophilus* α-amylase. A large number of mutations have been identified by sequencing in the course of refining the methodology.

The three-dimensional structures of two α-amylases, one from *Aspergillus oryzae*, the so-called Taka-amylase A, and the other from pig pancreas, have been determined by X-ray crystallography (Matsuura *et al.*, 1984; Buisson *et al.*, 1988). Both α-amylases have three domains: A, B and C. Overall, their structures appear very similar, but some differences are observed in the structure of domain B and in the orientation of domain C relative to domain A (Buisson *et al.*, 1988). The N-terminal domain A corresponds to the well-known $(\alpha/\beta)_8$ barrel motif, which was first seen in triose phosphate isomerase (Phillips *et al.*, 1978). In the middle of domain A, the chain loops out to make domain B, which forms a lid above the barrel. The active site is located at the top of the barrel on the C-terminal side of the parallel β-sheet. Substrate analogues bind in this site (Matsuura *et al.*, 1984; Buisson *et al.*, 1988), and there are three carboxylic acids around this site which are conserved in all known α-amylase sequences. In *B.stearothermophilus* α-amylase they correspond to residues D234, E264 and D331. Matsuura *et al.* (1984) propose that residues E231 and D297 of Taka-amylase A, corresponding to D331 and E264 of *B.stearothermophilus*, are the catalytic residues analogously to the mechanism in lysozyme. Buisson *et al.* (1988) propose that residues D196 and D300 of porcine pancreatic α-amylase, corresponding to D234 and D331 of *B.stearothermophilus*, are catalytic. A binding site for calcium, which is essential for activity (Steer and Levitzky, 1973), is located between domains A and B. The C-terminal region is folded into an antiparallel β-barrel and forms domain C. Although it has proved possible to crystallize the enzyme from *B.stearothermophilus* (Ogasahara *et al.*, 1970), X-ray structures for the bacterial α-amylases are not yet available.

It has been shown that tertiary structures are better conserved in evolution than amino acid sequences (Chothia and Lesk, 1986). Computer-aided molecular modelling has been used in a number of cases to construct a model of a protein of unknown structure which is related to a protein with a known structure [for a review see Blundell *et al.* (1987)]. α-Amylase genes from a variety of sources have been sequenced. Limited inter-species homology has been reported to be located in short stretches around the active site (Svensson, 1988). Using a sensitive method, we have found sufficient homology over almost all of the amino acid sequence which justifies homology modelling of the *B.stearothermophilus* sequence onto the Taka-amylase A structure. It is known that the three-dimensional structures of porcine pancreatic α-amylase and Taka-amylase A are similar. We show that the sequence homology between *B.stearothermophilus* and Taka-amylase A is more significant than between the former pair. The Taka-amylase A structure was chosen because it is currently the only publicly available α-amylase structure.

In this paper, we present a three-dimensional structural model of *B.stearothermophilus* α-amylase based on the structure of Taka-amylase A and show that the properties of the random mutants can be consistently interpreted with the aid of this model.

Materials and methods

Generation of point mutations

The α-amylase gene from *B.stearothermophilus* ATCC 12980 was cloned as a 1.9 kb long fragment to the phage vectors M13mp18 (coding strand) and M13mp19 (non-coding strand). The insert contains the complete coding region of the α-amylase gene (1650 bp). Random enzymatic mutagenesis of the cloned DNA was performed according to the method of Lehtovaara *et al.* (1988). The method can be summarized as follows. Starting from oligonucleotide primers, a population of molecules is synthesized under conditions where one nucleotide is limiting. This gives a population of molecules each of which ends just

before the omitted base. After removal of all nucleotides, misincorporation with reverse transcriptase is carried out under optimized conditions to give, in principle, a 1:1:1 ratio of the incorrect bases. Following synthesis of the second strand, the mutated dsDNA is transfected into *Escherichia coli*. The mutant library was generated in the course of developing the random mutagenesis method. Five oligonucleotides (20-mers) at ~200 nucleotide intervals were used to mutagenize the A nucleotides in the coding region of α-amylase in the M13mp19 construction. With one of these primers the template nucleotides G and C were also mutagenized. Only one primer was used to mutagenize T and G nucleotides in the M13mp18 construction (covering amino acids H208–L248).

Characterization of mutants

A suitable *dut*$^+$*ung*$^+$ strain of *E.coli*, such as JM109 or TG2, was transfected with the mutagenized DNA and the cells were plated on glucose minimal medium/H-top amylopectin azure plates. Expression of functional recombinant α-amylase in *E.coli* was detected as halos around the plaques. A three-grade scale was used to assay the α-amylase activity as follows: ++ for wild-type-size halos, + for smaller but visible halos and − for non-detectable halos. Mutations were identified by DNA sequencing.

Atomic coordinates for the starting model

The atomic coordinates for the starting model were from the structure of Taka-amylase A at 3.0 Å resolution [Matsuura *et al.*, 1984; available from the Protein Data Bank (Bernstein *et al.*, 1977) as entry 2TAA].

Sequence alignment

Because of a high level of noise, automatic Needleman–Wunsch (1970) procedures are not applicable to the alignment of sequences that are only ~20% identical, which is the case for most sequence pairs in our sample. The sensitive alignment program of Argos (1987) was used for initial pairwise alignments. Discrepancies between different pairwise alignments were resolved by visual pattern matching to produce a consensus alignment of the sequences. The positions of α helices and β strands were taken for pig pancreatic α-amylase from Buisson *et al.* (1988), and for Taka-amylase A they were calculated from the coordinates using the DSSP program (Kabsch and Sander, 1983). Helices and strands of pig pancreatic α-amylase and Taka-amylase A were matched. Gaps were introduced to maximize similarity across all sequences according to the following conservative substitution groups: {P, G}, {S, T}, {E, D, N, Q}, {H, Y, F, W}, {A, I, L, V, M, C}, {K, R}. At positions corresponding to a known helix or strand, however, gaps were not allowed in any sequence. The significance of pairwise alignments was assessed by the correlation of five physical characteristics of amino acids (hydrophobicity, bulkiness, turn preference, strand preference, refractivity index) as given by Argos (1987). This measure proved quite sensitive to small changes such as the movement of just a single residue in a gap. Argos (1987) recommends that alignments where the average correlation is at >2.0 SD above random be considered seriously. Adjustments were made to the alignment until the value was about this threshold for the majority of pairwise alignments.

Model building

It was assumed that the backbone structure in aligned positions is the same between Taka-amylase A and *B.stearothermophilus* α-amylase. Loops involving gaps were patched using fragments selected from the database of known structures (Jones and Thirup, 1986). One to three C(α) atoms at both ends of the loop were matched to the template (derived from Taka-amylase A). Different anchors were tried out until a fragment was found where the fit to C(α)s was good (deviation of superposed fragment to anchor atoms <1 Å). Ideally, all main chain atoms of the fragment and of the starting structure would overlap, but in practice the overlap could be restricted to just one residue at both ends. Main chain coordinates were replaced by those of the loop starting from the second residue in the N-side anchor region and ending with the second-last residue in the C-side anchor region. Graphics was used to make sure that the selected loops packed tightly to the rest of the structure. Structures that would clash with the rest of the molecule were rejected already when they were collected from the database. Side chain coordinates for aligned identical residues were copied from the Taka-amylase A structure, otherwise they were inserted in standard conformation. The conformation of F346 and W492 was changed to another staggered conformation because they initially clashed with the main chain.

Strain introduced to the model through sidechain substitutions and joining of loops was relaxed by the program CHARMm (Brooks *et al.*, 1983) using an adapted basis Newton–Raphson energy minimization with a distance-dependent dielectric constant and an 8 Å cut off for non-bonded interactions. The calcium ion was not included in the minimization.

A chain of six glucose units joined by α-1,4-linkages was built in the interactive molecular modelling program Quanta (Polygen Inc.). The dihedral angles in the aldose linkage were +90 and −31° for O5-C1-O4-C4 and C1-O4-C4-C3 respectively. The interaction energy of a water probe to protein was calculated in Quanta. Energy countours at an arbitrarily chosen level of −2.0 kcal/mol formed a tube which passed by the catalytic residues, and the glucose chain was fitted manually to it. Rotation of one glycosidic bond was necessary to fit the chain in contact with the catalytic residues (D234, E264, D331). The substrate was relaxed by energy minimization keeping the protein part frozen.

Results

Alignment of α-amylase sequences

The alignment of distantly related α-amylase sequences is far from obvious because only six blocks of residues are clearly conserved in them (Svensson, 1988). However, when the correlation of physical properties of amino acids is used as a measure, an acceptable alignment in terms of standard deviations above random sequence comparison scores can be made over the whole length of the sequences (Table I). The alignment in Figure 1 shows that the best conserved regions are in domains A and C, and these are segments around the catalytic residues and secondary structure elements where particularly hydrophobic patches are conserved. Although domain C lacks invariant residues, the patterns of alternating hydrophobic and hydrophilic residues are preserved in the β strands.

In the aligned α-amylase sequences (Figure 1) there are 31 positions where the nature of the amino acid is conserved. Of these, 13 positions require hydrophobic residues, three are invariantly glycine, three require an aromatic residue and seven require an invariant charged residue.

In domain B there is very little amino acid homology. Further, the arrangement of β strands in the two known structures, porcine pancreatic α-amylase and Taka-amylase A, is distinctly different (Buisson *et al.*, 1988). In addition, in the amyloliquefying

Table I. Quality of the alignment

Sequence		1	2	3	4	5	6	7	8	9	10
1	*S.hygroscopicus*		8.4	3.9	1.4	0.7	3.0	0.3	1.3	1.4	−0.1
2	Pig pancreas	38		5.0	1.2	1.4	2.8	1.6	2.1	2.3	1.7
3	*B.subtilis*	24	24		2.3	2.1	2.7	0.5	1.5	2.1	1 8
4	*B.circulans*	21	18	21		4.4	5.5	1.3	1.3	1.5	0.8
5	*S.fibuligera*	19	10	21	28		14.5	3.0	3.3	3.5	1.1
6	Taka-amylase	22	23	20	27	52		3.7	3.2	3.4	1.0
7	*B.stearothermophilus*	20	16	19	22	23	24		17.7	17.1	2.2
8	*B.licheniformis*	18	16	21	20	23	21	66		20.3	3.2
9	*B.amyloliquefaciens*	19	17	21	21	23	23	66	81		3.7
10	Barley	17	17	18	18	19	20	24	25	26	

Upper part of the table: number of SD above random for the alignment in Figure 1 over five parameters as given by the sensitive homology program of Argos (1987). If 2.5 standard deviations is used as a criterion, all but barley are satisfactorily aligned with Taka-amylase A. The structure of Taka-amylase A (*Aspergillus oryzae*) was used as a template for the structural model of *B.stearothermophilus*. Lower part of the table: percentage of identical amino acids of aligned positions, i.e. positions not occupied by a gap in either sequence of a pairwise comparison. Aligned positions account for at least 70% of the length of any sequence in any pairwise comparison.

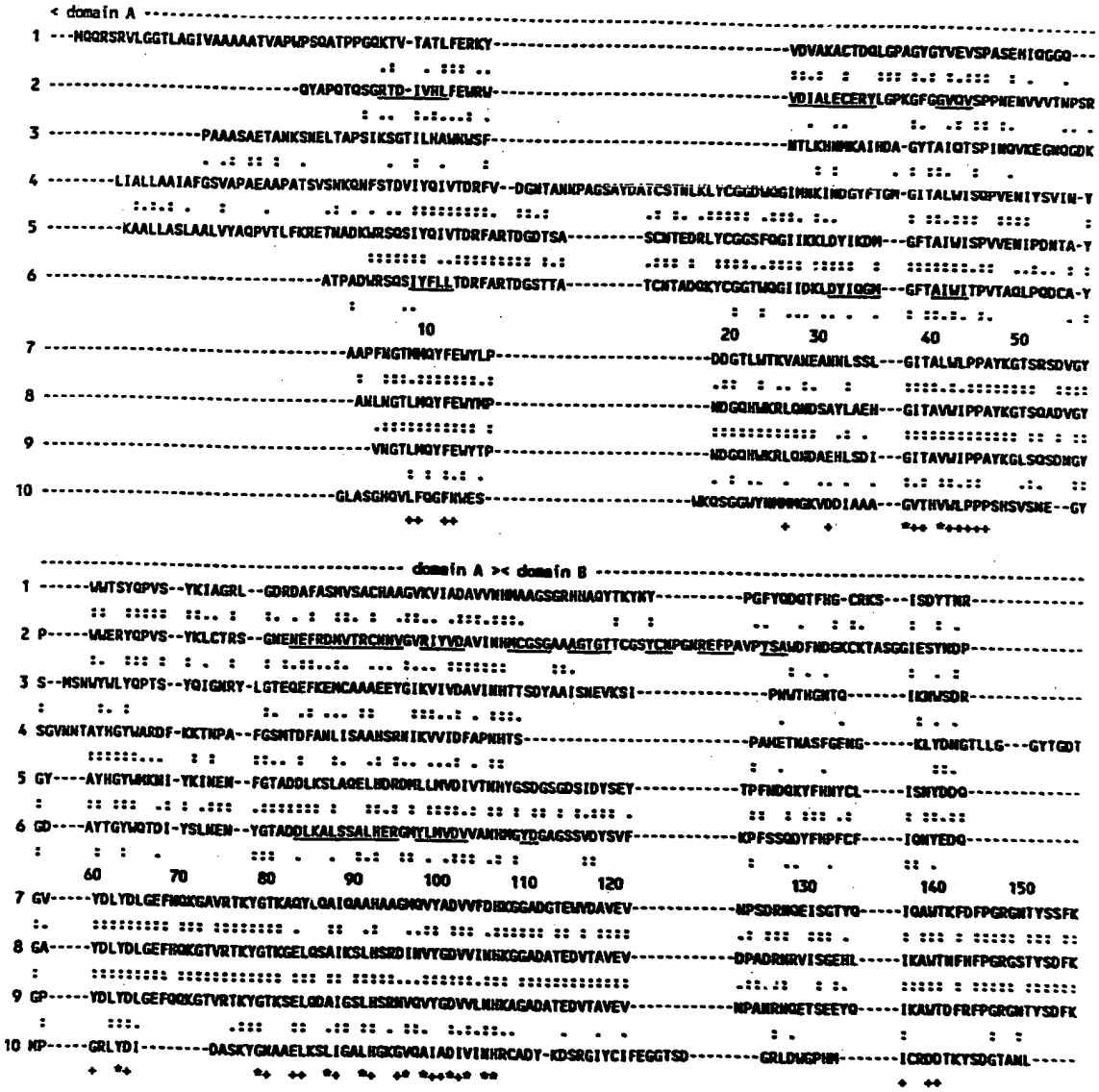

Fig. 1.

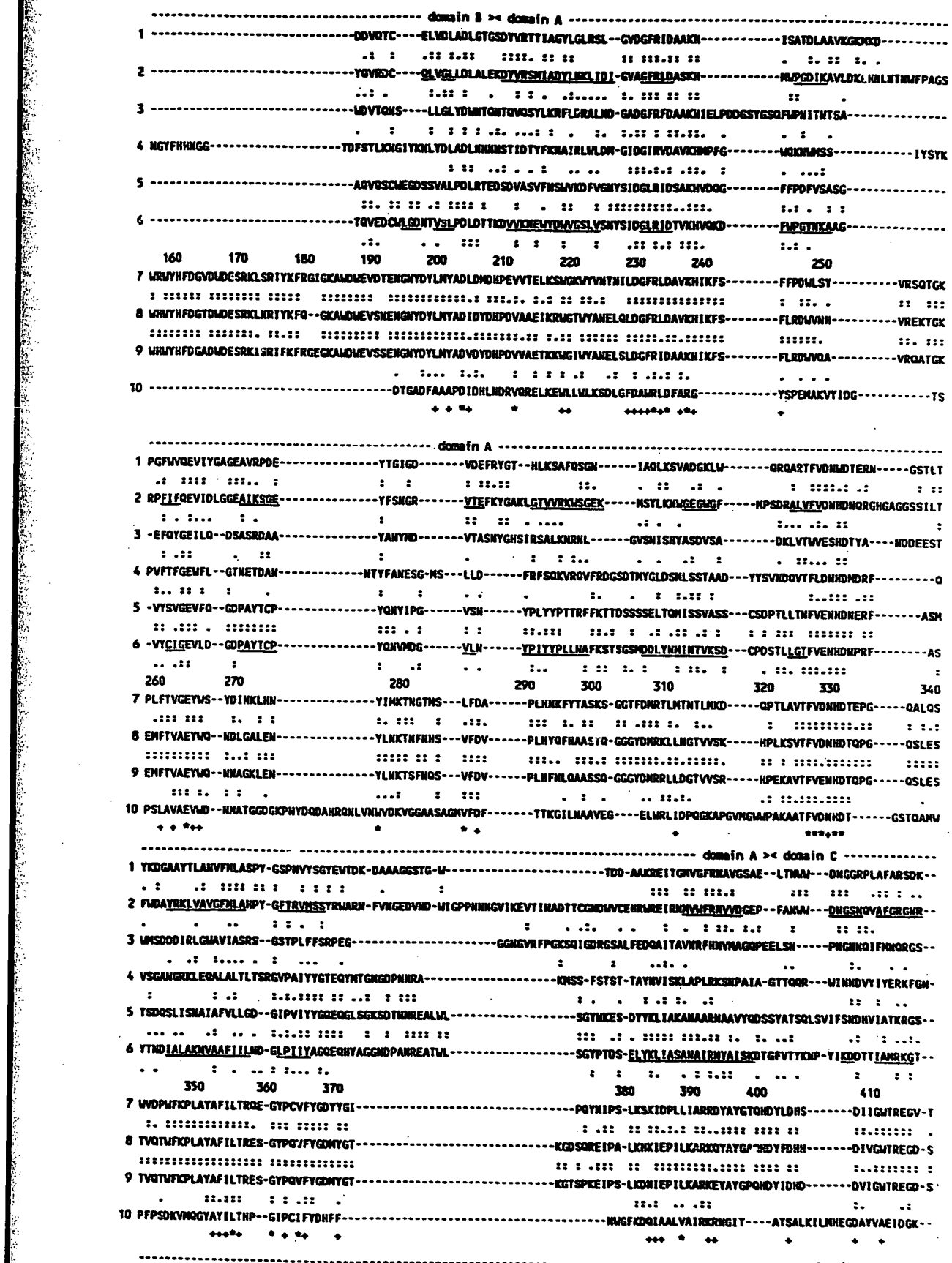

Fig. 1 (continued).

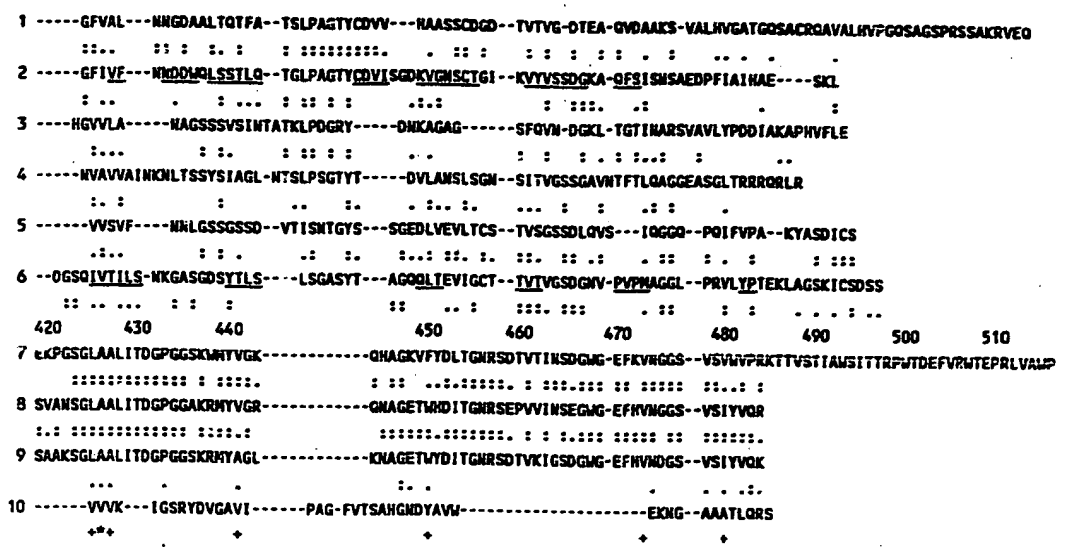

Fig. 1. Sequence alignment of α-amylases from different sources. Aligned α-amylase amino acid sequences from: 1, *Streptomyces hygroscopicus* SF-1024 strain AA69-4 (Hoshiko *et al.*, 1987); 2, pig pancreas (Pasero *et al.*, 1986); 3, *Bacillus subtilis* strain N7 (Yamane *et al.*, 1984); 4, *B.circulans* (Nishizawa *et al.*, 1987); 5, *Saccharomyces fibuligera* (Yamashita *et al.*, 1985); 6, *Aspergillus oryzae* (Matsuura *et al.*, 1984); 7, *B.stearothermophilus*; 8, *B.licheniformis* (Yuuki *et al.*, 1985); 9, *B.amyloliquefaciens* (Takkinen *et al.*, 1983); 10, barley (Rogers, 1985). Five more mammalian α-amylase sequences are included in NBRF protein database (Barker *et al.*, 1988). They have >80% identical residues with the pig pancreatic α-amylase and were therefore excluded from the alignment. Identical amino acids in adjacent sequences are marked by a colon. A dot marks changes within the following groups: {P, G}, {S, T}, {E, D, N, Q}, {H, Y, F, W}, {A, I, L, V, M, C}, {K, R}. Gaps are denoted by −. An asterisk below the alignment marks positions conserved in all sequences, and a plus denotes positions where variability is limited to two different groups. α Helices and β strands are underlined in the known structures. Gaps correspond to surface loops in the structure of Taka-amylase A. Residue numbers for *B.stearothermophilus* are shown above its sequence.

Bacillus α-amylases, domain B contains a 43−45 residues long insert relative to Taka-amylase A. For these reasons the proposed structure for domain B should be treated with considerable caution.

Model building

The alignment shown in Figure 1 was used to construct a three-dimensional model of *B.stearothermophilus* α-amylase using the structure of Taka-amylase A as a template (Figure 2). Due to the lack of a suitable structural template, two segments have been omitted from the model: the 45-residue insert in domain B—the model joins residue 142 to residue 189—and the C-terminal extension of 19 residues (residues 497−515). Without these two segments, construction of the structural models was relatively straightforward since the remaining shorter insertions and deletions involve only a few residues. Insertions or deletions also occur in places where the polypeptide chain folds back on itself so that loops of different lengths are easy to accommodate. Thus the structural changes made during modelling were minimized. One difficult region was between amino acids 444 and 456, in which a loop between two β strands in the Taka-amylase A structure is shortened by five residues. This deletion could only be incorporated by using a long extended loop which caused concomitant restructuring of this region by displacing one β strand. As a result, domain C appears more flat and open in our model than in Taka-amylase A.

Mutant library

Point mutations distributed throughout the α-amylase gene were generated by an enzymatic misincorporation method (Lehtovaara *et al.*, 1988). In the work presented here, mutants were identified by sequencing and subsequently assayed on starch plates for enzyme activity. In this sample of 98 phenotypic mutants, there are 31 which show no activity, 28 with decreased activity and 39 with wild-type enzymatic activity (Table II). Seventy-five different residues are hit by at least one phenotypic mutation,

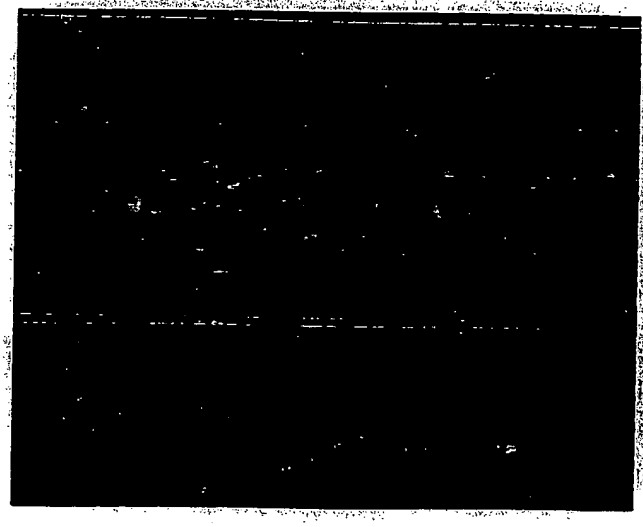

Fig. 2. Structural model of *B.stearothermophilus* α-amylase based on the known structure of Taka-amylase A. Left: Taka-amylase A, right: model of *B.stearothermophilus* α-amylase. Only the N, C(α) and C atoms of the main chain are shown. The model was constructed based on sequence homology to Taka-amylase A, the structure of which has been determined by X-ray crystallography. A 45-residue insert from the back of domain B and a 17-residue extension to the C terminus have been omitted. The light blue sphere shows the location of the essential calcium ion as found in Taka-amylase A. Red, green and yellow spheres mark the probable catalytic residues. Loops 17−38 and 333−350 of Taka-amylase A and the corresponding loops 15−19 and 370−374 of *B.stearothermophilus* α-amylase are coloured yellow. Domain B is coloured orange and domain C green in *B.stearothermophilus* α-amylase. Domain A (blue) is an $(\alpha/\beta)_8$ barrel.

and Figure 3 shows that they are distributed over most of the protein. Clusters of residues at which mutations affect activity are seen around the active site, and at the interface between

Table II. Mutants generated by the random mutagenesis method

					Activity	Observed amino acids
(i) Mutations to stop codons						
K48stop					–	
R126H	**E129stop(S131S)**				–	
E129stop					–	
K141stop					–	
K155stop					–	
K257stop					–	
Y439F	**K442stop**				–	
(ii) Mutations in the active site						
R232W					–	232: R
D234G					–	234: D
H238N					+	238: HG
(V236V)	**H238Y**				–	
H238Y					–	
(L259L)	**T261A**				++	
T261S	**E264V**				–	261: WIQTCSA
E264D	**Y265S**				–	264: E
Y265S					++	265: YVWI
Y265F					+	
D331A					–	331: D
T332P	**E333D**				–	332: TNM
E333A					++	333: EQYP gap
(iii) Mutations in the substrate-binding groove						
Y15F	***D18V***				++	15: YERSK
Y15F	***D18A***				–	18: YDNQ gap
Y15F	***D18V***	**D19V**	**T21S**	**L22F**	+	19: DSC gap
Y15S	***D18V***	**D19V**	**T21P**	**L22F**	+	21: TQG gap
Y15S	***D18A***	**D19V**	**T21P**	**L22F**	+	22: LHWSTD gap
D18V	**D19V**				+	
D18V	**D19A**				+	
D18A	**D19V**	**T21S**			++	
D18A	**D19V**	**T21P**	**L22F**		+	
S51R	***S53R***				–	51:SPEYVQIA
D54V	**(V55V)**	***Y57C***			–	53: SAVDNTGH
D54A	**(V55V)**	***Y57S***	**(G58G)**		–	54: DCTIQNS gap
D54V	**(V55V)**	***Y57S***	**(G58G)**		–	57: RKY gap
Y60F	***D61V***	***Y63F***			–	60: YGW
Y63C					+	61: DRTHWE
K237R					++	63: Y
P334S					++	
Q336L					+	336: QTDSGA gap
Q336H					++	
Q336P					+	
Y369S					+	369: YMFEPR
Y370S					+	370: YFQEWR
(iv) Mutations in or around the interface between domains A and C						
D343A					+	343: DQPNGS
K347N					+	347: KVLIRY
Y351S					+	351: YIVQWN
I354L					+	354: IFVAM
R413G					–	413: RIKQ
R413W	**E414A**				–	414: EDKRGS
R413W	**E414V**				–	
E414V					++	
E414D					++	417: TS gap
T417P	***E418A***				+	418: ES gap
(A426A)	**I428F**				–	428: ILFA gap
S435F					–	435: SAYLGW
S435C	***K436E***				+	436: KDSQAFT
(v) Mutations in other buried sites (solvent-accessible surface of wild-type residue <50 Å^2)						
E13V					++	
D19V	**T21S**				++	

Table II (continued).

				Activity	Observed amino acids
D19V	T21P	(L22L)	T24P	++	
E29D				+	29: EDKA
N32I	L33F	S35C		+	32: NYHDMC
L33F				−	33: LIGKTE
L33F	S35C			+	35: SEDAGFIYQ
K48T				++	
E129G				++	
Q254P				++	
I277F				++	
L291F				+	291: LTIFS
G335C				−	335: GFN gap
G335C	A337T			++	337: ASQE
A337T				++	
(vi) Mutations in other exposed sites (solvent-accessible surface of wild-type residue >50 Å^2					
L22F				++	
K25N				−	25: KRMG gap
N31Y				+	31: NAED
N127T				+	127: NLQTWDF
T133S	Y134F			++	
Y134S				++	
Y134F				++	
Q135L				++	
Q137P				++	
A138T				++	
N149I				++	
H208L				++	
T213A				++	
K216E				++	
S217G				++	
S217C				++	
K220R				++	
T225P				+	225: TEDNMI gap
Y250F				+	250: YHAVST
T255P				−	255: TSA gap
K257T				++	
K257Q				++	
K257M				++	
K419N				++	
K419T				−	419: KVA gap
Y439F	K442Q			+	439: YAGIVSQLE
K442N				++	442: KRLSDNQA gap
Q443H				++	

The wild-type amino acid is given on the left and the amino acid it is mutated to is given on the right of the residue number. Deleterious substitutions are printed in **bold**. Note that some of the mutants have multiple substitutions; comparisons with corresponding single mutants sometimes suggest the causative change. In clusters (like the one around residue 20), where the causative mutations are less clear, the primary suspects are printed in *italics*. Consequently, all other mutations have been interpreted as being neutral. Silent mutations seen only at the DNA level are in parentheses. Enzyme activities of the mutant clones were estimated on starch plates as compared to the wild-type clone: ++ wild-type activity, + less than wild-type but greater than zero, − zero activity. No clone showed drastically elevated activity. Mutants are listed grouped according to their location in the structural model. The solvent-accessible surface area was calculated using the program DSSP (Kabsch and Sander, 1983) from the model of wild-type *B.stearothermophilus* α-amylase. Residues 1−104 and 208−400 belong to domain A, residues 105−207 to domain B and residues 401−515 to domain C. For convenience of comparison to other α-amylase sequences, the right-most column lists for + and − mutants which amino acids are observed in natural α-amylases at given positions. The data for this listing also includes five mammalian sequences (PIR codes ALHUS, ALHUP, ALMSS, ALMSP, ALRTP) which are highly homologous to pig pancreatic α-amylase and which were excluded from Figure 1.

domains A and C. The relationship between the effects of different mutations and their location on the model is discussed in more detail below.

Active site

There are many conserved and buried charged residues around the active site. D331 is relatively accessible to solvent but the other two proposed catalytic residues, D234 and E264, are close neighbours and appear almost buried at the bottom of a pocket. The most striking mutation in this region is the chemically homologous substitution E264D, which is inactive (in a double mutant with Y265S, which as a single mutation has wild-type activity). This strongly suggests that E264 is essential for catalysis. In addition the mutation D331A on one hand and the mutant D234G on the other are inactive, suggesting that the carboxyl groups of all three, D234, E264 and D331, play a role in the enzymatic mechanism. In the energy minimized model, R232 is in the centre of a network of hydrogen bonds and salt bridges, which may fix the catalytic residues in the correct orientation for catalysis. R232 is connected to D234, D101, H330

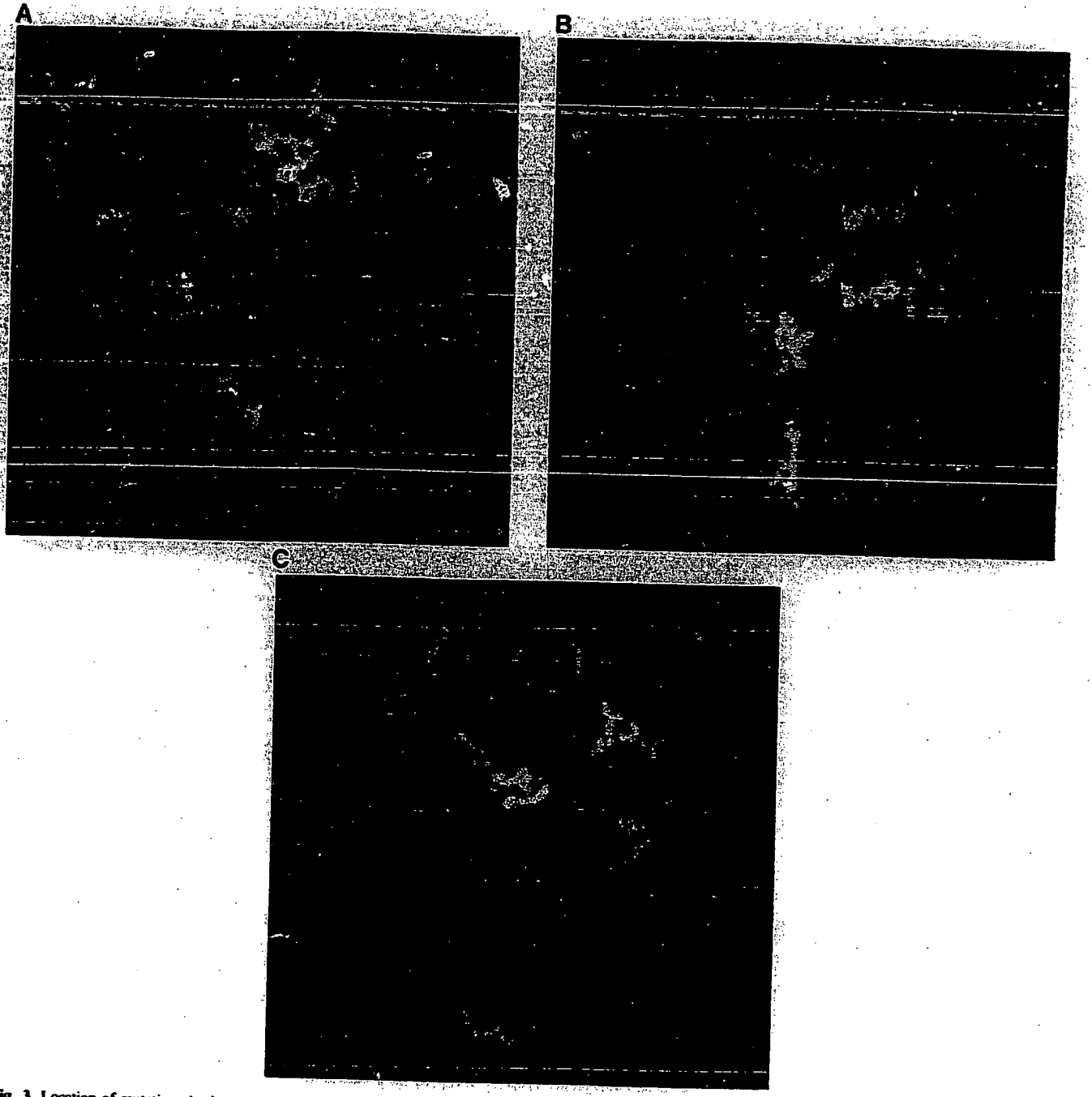

Fig. 3. Location of mutations in the structural model of *B.stearothermophilus* α-amylase. Residues which upon mutation produce (A) loss of activity (− mutants), (B) reduced activity (+ mutants) or (C) wild-type activity (+ + mutants). The side chain atoms are drawn as spheres with 0.5 times their van der Waals radius (atom colouring: C green, N blue, O red). Most − and + mutants are clustered around the active site and the interface between domains A and C.

and E264, and further to W42, D331, N329 and D288. All of these residues are highly conserved. In other α-amylases, W42 can be replaced by Q, N329 by S and D288 by N or E or A but all the other mentioned residues are invariant (Figure 1). The mutants R232G and R232W are inactive, thus demonstrating the functional importance of this conserved arginine residue.

Substrate-binding groove

In the model there is a clearly visible groove winding across the top of the α/β barrel in domain A, into which a six-unit amylose chain was docked. Because D234 and E264 are at the bottom of a deep pocket, the long substrate chain could only be brought into contact with them by introducing a sharp bend in the chain. The glucose units make extensive contacts with a number of residues equivalent to those proposed for Taka-amylase A (Matsuura *et al.*, 1984) and porcine pancreatic α-amylase (Buisson *et al.*, 1988). Those residues that are in contact with the substrate at the subsites adjacent to the hydrolyzed bond are highly conserved in all α-amylases, but more variation is allowed at the other subsites.

In Taka-amylase A the region of the substrate-binding groove furthest from the catalytic site is blocked by the adjacent protruding loops 17–38 and 333–350 (residue numbers according to the Taka-amylase A sequence). They correspond to the much shorter loops 15–19 and 370–374 in *B.stearothermophilus* α-amylase (Figure 2). We speculate that the 'missing' subsites might explain the fact that Taka-amylase A is more saccharifying than *Bacillus* α-amylases (Priest, 1984) by biasing the substrate specificity of Taka-amylase A towards residues at the end of the amylose chain.

In the model of *B.stearothermophilus* α-amylase there are notably many tyrosine side chains lining the substrate-binding groove. Several mutations, single or multiple, in these residues (Y15, Y57, Y60, Y63, Y369, Y370) lead to decreased activity (Table II). From its location and mutant data (Table II), Q336 is also implied in substrate binding in *B.stearothermophilus* α-amylase, though it has no counterpart in Taka-amylase A, for example (Figure 1). W266 is in contact with the docked substrate. It corresponds to tryptophan 206 in the barley α-amylase, which has been pinpointed as essential for enzymic function in biochemical mapping (Gibson and Svensson, 1987). Interestingly, in some α-amylases this position is occupied by hydrophobic residues other than tryptophan.

Interface between domains A and C

A second major cluster of inactivating mutations is found in and around the interface between domains A and C, ~30 Å away from the catalytic site (Table II and Figure 3). In low solution X-ray studies of porcine pancreatic α-amylase a second binding site for substrate analogues has been identified near the N-terminus of the eighth helix in domain A that is facing domain C (Payan *et al.*, 1980). The function of this site is unknown, but it has been implied that it may play a role either in anchoring the enzyme to starch or in the regulation of activity (Buisson *et al.*, 1981). In our model both salt bridges and hydrophobic contacts are found across the interface between domains A and C. The latter interactions are affected in the inactive mutant I428F, which is located on a β strand in domain C, and in the mutant I354L, which is on the seventh α helix in domain A and has much decreased activity although this amino acid substitution is conservative. The dramatic effects of such 'innocent-looking' mutations seem to imply that the exact orientation of domain C to domain A is crucial for activity. In the model of *B.stearothermophilus* α-amylase these two residues are opposite to each other. Many combinations of hydrophobic residues are found at these positions in the other α-amylases (Table II, Figure 1). Since, particularly in domain C, many gaps were introduced into the alignment (Figure 1), the structural details of domain C and its contacts with domain A are obviously different in the different enzymes. This is clearly illustrated by comparing the known structures of Taka-amylase A and porcine pancreatic α-amylase (Buisson *et al.*, 1988). That domain C is important for activity is also shown by the effect of a mutation that causes a premature stop codon at residue 442. This removes roughly two-thirds of domain C and leads to inactivation of the enzyme.

Calcium-binding site

The structure of the calcium binding site between domains A and B is similar in the two known structures, and the invariant D203 is one of the ligands of the calcium ion (Buisson *et al.*, 1988). D105 and the main chain carbonyl oxygen from the invariant H238 are other ligands. D105 is replaced by an asparagine in the other α-amylases (Figure 1). The importance of H-bonds involving the sidechain of H238 is suggested by the mutations H238N, which is still functional, and H238Y, which is inactive (Table II).

Buried sites are typically more sensitive to mutation than exposed sites

The great majority of amino acid replacements which do not much affect the activity of the enzyme are located at exposed sites in the 3-D model. Mutations affecting the active site region and the A–C interdomain region which inactivate α-amylase were described above. Several other examples, where even conservative mutations are deleterious, were found at buried sites. For example, the inactive mutant L33F and the mutant L291F, which has reduced activity, are located in the interior of domain A. Interestingly, the substitution corresponding to L291F is found in *B.circulans* α-amylase. An examination of the structure of Taka-amylase A and models of *B.circulans* and *B.stearothermophilus* α-amylase indicated that Phe can be accommodated at this position in the core of *B.circulans* α-amylase due to a number of compensating substitutions to smaller residues in its neighbourhood, but it could not be fitted to the other two structures without steric clashes.

E29D, located at a partially buried site, has reduced activity. In the three-dimensional model E29 forms a salt bridge to K25. This seems to be an important interaction in the *B.stearothermophilus* enzyme, since the mutant K25N is inactive. The other liquefying *Bacillus* enzymes have D at position 29, but the reduction in size of this residue is compensated by R at position 25. No corresponding salt bridge is present in the structure of Taka-amylase A.

At certain sites proline may be accommodated without effects on the activity, and at other sites it cannot be accommodated (Table II). For example, the mutations Q254P, leading to fully active enzyme, and T255P, leading to inactive enzyme, are in neighbouring residues in a surface loop.

Interactive effects in multiple mutations

Table II shows that generally the effects of amino acid substitutions are additive. There are, however, a number of exceptions. For example, G335C alone gives an inactive enzyme, but activity is recovered if A337T is also present. A337T alone has no effect on activity. Other interesting examples include the clusters of mutations around residues D18 and L33.

Discussion

We describe here the construction of a molecular model of *B.stearothermophilus* α-amylase using a novel alignment of amino acid sequences from many distantly related α-amylases and the available information on secondary/tertiary structures. We also present the phenotype and spatial location of 98 separate randomly generated amino acid mutants and note that inactivating mutations are clustered in the model, revealing functionally important regions. Thus, the model provides a rationale for interpretation of the mutations, and *vice versa*, data on the mutants provides evidence supporting the model.

Porcine pancreatic α-amylase and Taka-amylase A have only 23% identical amino acids (Table I) but very similar three-dimensional structures. We have assumed that all α-amylases, which have typically ~20% identical residues, share this common structural design. There are several well known examples of protein families, such as the globins and the trypsin-like serine proteases, where the sequences can diverge even more but the three-dimensional structures are well conserved. The correct alignment of the amino acid sequences to identify residues

occupying topologically equivalent positions is critical to the success of homology modelling (Read *et al.*, 1984). Argos (1987) has described a cautionary example concerning the pairwise alignment of lactate and malate dehydrogenase: the sequence alignment optimal in terms of amino acid identity does not coincide with the alignment derived from the optimal superposition of the tertiary structures. However, the alignment of α-amylases described here is supported by structural information concerning calcium ligands and secondary structure. In addition, amino acids of similar physical properties are well aligned. At least 70% of residues in all sequence pairs are aligned in Figure 1 and, statistically, the alignment is significant (Table I). It is generally accepted that if the sequence of a protein with unknown tertiary structure can be related to one with known architecture, then the main-chain folding of the latter provides an adequate base for structural modelling (Blundell *et al.*, 1987). However, in the course of evolution, mutations are accommodated in protein structures through small shifts in the relative position of secondary structure elements (Chothia and Lesk, 1982a). Such effects cannot be accounted for in modelling procedures which use a rigid framework. Refinement by energy minimization results only in small shifts in relative atomic positions, typically in the order of 1 Å at most. Thus, it is clear that this model represents only one possible low-energy conformation of the protein. Despite these reservations, we were pleased to note that the α-amylase model is consistent with biochemical data obtained from random mutagenesis.

For this work, it was more important to gather information on a large number of mutants than to analyze a few in detail. Mutants generated by random mutagenesis were identified by sequencing, since it was important to understand what types of base changes the misincorporation procedure produced (Lehtovaara *et al.*, 1988). Around 40% of phenotypic mutants in the sample described here retained wild-type activity. The remainder had reduced activity or were inactive.

The loss of enzyme activity observed in the plate assay could be assayed by: (i) impaired function of secreted enzyme or (ii) a reduction in the amount of enzyme due to destabilization of the secreted protein or impaired folding and even degradation inside *E.coli*. Mutational analysis of a number of proteins has shown that amino acids which are critical to protein stability are generally relatively rigid and inaccessible to solvent in the folded protein whilst substitutions at mobile and exposed sites usually have little effect on protein stability (reviewed by Alber, 1989). Exceptions are made by the introduction of prolines or replacement of conformationally special glycines, which can propagate shifts in the folded structure even if the mutation is at an exposed site (Alber, 1989). Our interpretations concerning the effects of the mutations are based on the location and interactions of the mutated residues in the structural model. Mutations which apparently impair function are clustered together in three-dimensional space at two surface sites: the active site cleft and the domain A–domain C interface. In the substrate-binding groove, a number of exposed residues, which may form hydrogen bonds to the substrate, were shown to be sensitive to mutation. The catalytic center and calcium-binding site are also identified from the known X-ray structures and by sequence conservation. Seven absolutely conserved charged residues in the family of α-amylases are located here. Four of these residues were mutated in our sample and led to inactive enzyme in each case. Apart from the clusters of deleterious mutations mentioned above, exposed residues were found to be less sensitive to mutations than buried ones. A number of mutations to proline and mutations substituting a larger sidechain at buried sites were apparently deleterious because they disrupt the tertiary structure.

The pH dependence of α-amylase suggests a lysozyme-type mechanism employing two carboxylic acids, one ionized and the other protonated in the active form. The adjacent E264 and D234 are buried, and therefore may have elevated pK_as, whereas D331 is more exposed and is expected to have a normal pK_a. Matsuura *et al.* (1984) have argued that in Taka-amylase A two carboxyl groups in different environments are required for activity and therefore the residue corresponding by homology to D331 in *B.stearothermophilus* α-amylase is a catalytic residue. The proposals by Matsuura *et al.* (1984) and Buisson *et al.* (1988) agree on the role of D331 as a catalytic residue in α-amylases but differ with respect to the roles of the neighbouring residues E264 and D234. Buisson *et al.* (1988) reason that D234, which has more direct interactions with the calcium-binding site, is likely to be the other catalytic residue. However, our mutant data very strongly suggest that E264 is a catalytic residue.

Domain C is shown here to be indispensable for activity since a cluster of mutations far from the active site reduce enzyme activity. Subtle conformational changes at the interface between domains A and C, which we assume accompany the mutations I428F and I354L, seem to have a large effect on activity. It has been proposed for mammalian α-amylases that a site at the domain A–domain C interface might have a regulatory role in that substrate analogues bind to this region and affect activity (Payan *et al.*, 1980; Buisson *et al.*, 1981). It is not clear why a bacterial exoenzyme should require this type of regulation. We therefore propose that domain C plays an important role in starch hydrolysis by orientating the active site cleft of domain A correctly with respect to the amylose chain.

In conclusion, we suggest that this approach of integrating computer-aided molecular modelling to the rationalization of random mutagenesis data can contribute substantially to an understanding of protein structure and function, even in cases like this where the crystal structure of that protein is not available.

Acknowledgements

We thank Chris Sander for a helpful comment and Matti Saraste for critical reading of the manuscript. We are grateful to Tuula Teeri and Leif Laaksonen for their momentous input in providing for excellent molecular modelling facilities at VTT and FSCC. Expert technical assistance in making and analyzing the mutants by Kariitta Berg is gratefully acknowledged. This work was supported by the Academy of Finland (S.R.C.) and the Technology Development Centre (TEKES).

References

Alber,T. (1989) *Annu. Rev. Biochem.*, 58, 765–798.
Argos,P. (1987) *J. Mol. Biol.*, 193, 385–396.
Barker,W.C., Hunt,L.T., George,D.G., Yeh,L.S., Chen,H.R., Blomquist,M.C., Seibel-Ross,E.I., Elzanovski,A., Bair,J.K., Lewis,M.T., Davalos,D.P. and Ledley,R.S. (1988) National Biomedical Research Foundation, Georgetown University Medical Center, Washington DC.
Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, 112, 535–542.
Blundell,T.L., Sibanda,B.L., Sternberg,M.J.E. and Thornton,J.M. (1987) *Nature*, 326, 347–352.
Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) *J. Comp. Chem.*, 4, 187–217.
Buisson,G., Duee,E., Haser,R., Payan,F. and Darbon,N. (1981) *Acta Crystallogr.*, A37, C-34 (suppl.).
Buisson,G., Duée,E., Haser,R. and Payan,F. (1988) *EMBO J.*, 6, 3909–3916.
Chothia,C. and Lesk,A.M. (1982a) *J. Mol. Biol.*, 160, 309–323.
Chothia,C. and Lesk,A.M. (1982b) *J. Mol. Biol.*, 160, 325–342.
Chothia,C. and Lesk,A.M. (1986) *EMBO J.*, 5, 823–826.
Gibson,R.M. and Svensson,B. (1987) *Carlsberg Res. Commun.*, 52, 373–379.

Hoshiko,S., Makabe,O., Nojiri,C., Katsumata,K., Satoh,E. and Nagaoka,K. (1987) *J. Bacteriol.*, **169**, 1029–1036.
Jones,T.A. and Thirup,S. (1986) *EMBO J.*, **5**, 819–822.
Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.
Knowles,J.R. (1988) *Science*, **236**, 1252–1258.
Lehtovaara,P.M., Koivula,A.K., Bamford,J. and Knowles,J.K.C. (1988) *Protein Eng.*, **2**, 63–68.
Matsuura,Y., Kusunoki,M., Harada,W. and Kakudo,M. (1984) *J. Biochem.*, **95**, 697–702.
Needleman,S.B. and Wunsch,C.D. (1970) *J. Mol. Biol.*, **48**, 443–453.
Nishizawa,M., Ozawa,F. and Hishinuma,F. (1987) *DNA*, **6**, 255–265.
Ogasahara,K., Imanishi,A. and Isemura,T. (1970) *J. Biochem.*, **67**, 65–74.
Payan,F., Haser,R., Pierrot,M., Frey,M., Astier,J.P., Abadie,B., Duee,E. and Buisson,G. (1980) *Acta Crystallogr.*, **B36**, 416–421.
Pasero,L., Mazzi-Pierron,Y., Abadie,B., Chicheportiche,Y. and Marchis-Mouren,G. (1986) *Biochim. Biophys. Acta*, **869**, 147–157.
Phillips,D.C., Sternberg,M.J.E., Thornton,J.M. and Wilson,I.A. (1978) *J. Mol. Biol.*, **119**, 329–351.
Priest,F.G. (1984) In Cole,J.A., Knowles,C.J. and Schlessinger,D. (eds), *Aspects of Microbiology 9: Extracellular Enzymes*. Van Nostrand Reinhold UK, Wokingham, p. 34.
Read,R.J., Brayer,G.D., Jurasek,L. and James,M.N.G. (1984) *Biochemistry*, **23**, 6570–6575.
Rogers,J.C. (1985) *J. Biol. Chem.*, **260**, 3731–3738.
Steer,M.L. and Levitzky,A. (1973) *FEBS Lett.*, **31**, 89–92.
Svensson,B. (1988) *FEBS Lett.*, **230**, 72–76.
Takkinen,K., Pettersson,R.F., Kalkkinen,N., Palva,I., Söderlund,H. and Kääriäinen,L. (1983) *J. Biol. Chem.*, **258**, 1007–1013.
Yamane,K., Hirata,Y.I., Furusato,T., Yamazaki,H. and Nakayama,A. (1984) *J. Biochem.*, **96**, 1849–1858.
Yamashita,I., Itoh,T. and Fukui,S. (1985) *Agric. Biol. Chem.*, **49**, 3089.
Yuuki,T., Nomura,T., Tezuka,H., Tsuboi,A., Yamagata,H., Tsukagoshi,N. and Udaka,S. (1985) *J. Biochem.*, **98**, 1147–1156.

Received on July 10, 1989; revised on October 18, 1989

Site-Directed Mutagenesis of a Thermostable α-Amylase from *Bacillus stearothermophilus*: Putative Role of Three Conserved Residues[1]

Mauno Vihinen,*[,2] Pauli Ollikka,* Jukka Niskanen,* Peter Meyer,* Ilari Suominen,* Matti Karp,* Liisa Holm,** Jonathan Knowles,** and Pekka Mäntsälä*

*Department of Biochemistry, University of Turku, 20500 Turku, Finland; and **Biotechnical Laboratory, VTT, 02150 Espoo, Finland

Received for publication, August 16, 1989

The relationship between structure, activity, and stability of the thermostable *Bacillus stearothermophilus* α-amylase was studied by site-directed mutagenesis of the three most conserved residues. Mutation of His-238 to Asp involved in Ca^{2+} and substrate binding reduced the specific activity and thermal stability, but did not affect the pH and temperature optima. Replacement of Asp-331 by Glu in the active site caused almost total inactivation. Interestingly, in prolonged incubation this mutant enzyme showed an altered end-product profile by liberating only maltose and maltotriose. Conservative mutation of the conserved Arg-232 by Lys, for which no function has yet been proposed, resulted in lowered specific activity: around 12% of the parental enzyme. This mutant enzyme had a wider pH range but about the same temperature optimum and thermal stability as the wild-type enzyme. Results obtained with different mutants were interpreted by computer aided molecular modeling.

α-Amylase is widely produced by microbes, plants, and animals to hydrolyze starch and other linear and branched 1,4-glucosidic oligo- and polysaccharides. Almost all α-amylases studied (for a review see Ref. *1*) have a few conserved regions (*2*, *3*) with approximately the same length and spacing. One of these regions resembles the Ca^{2+} binding site of calmodulin and the others are thought to be necessary for the active center and/or binding the substrate (*2*). The refined three-dimensional structure of the *Aspergillus oryzae* α-amylase, called Taka-amylase A (TAA), has been determined at 3 Å resolution (*4*). The *A. oryzae* α-amylase has been suggested to bind seven glucose residues to its subsites on the surface of the protein and one of these subsites, containing amino acids Glu-230 and Asp-297, was proposed to be the catalytic center (*4*). Recent structural studies on porcine pancreatic α-amylase (PPA) revealed that Glu-230 is more likely to be involved in substrate binding (*5*). The catalytic amino acids of TAA are thus Asp-206 and Asp-297 (*5*). The calcium ion, found in virtually all α-amylases and required for stability (*1*), is bound to four amino acids in TAA and PPA (*5*).

We have used the cloned, thermostable *Bacillus stearothermophilus* α-amylase (BStA) in studies concerning expression and secretion (*6-9*). Here we studied the role of the three most conserved amino acids in maintaining the structure, activity, and stability of the thermostable α-amylase. The structure of the enzyme was modeled from the known structure of TAA, and this model was used to interpret the results obtained with mutated enzymes. Conserved residues Arg-232, His-238, and Asp-331 were all shown to have important functions.

MATERIALS AND METHODS

Strains, Culture Conditions, and Assays—Strains *Escherichia coli* JM103 (*10*); MC1061 (*11*); RZ1032 (*12*), *Bacillus subtilis* 1A40 (Bacillus Genetics Stock Center, Ohio), and phages M13mp18 and mp19 (*13*) have been described. L broth, 2xTY, and M9 mediums were supplemented with 100 μg/ml ampicillin or 10 μg/ml kanamycin when necessary (*14*). Plasmid pCSS6 (*7*) and p602/15HT have been described. α-Amylase activity was measured as described (*7*) and protein concentration was determined according to Lowry (*15*).

DNA Techniques and In Vitro Mutagenesis—Isolation of plasmids and single-stranded DNA sequencing was according to Maniatis *et al.* (*14*) and double stranded sequencing was according to Zagursky *et al.* (*16*). Competent *E. coli* (*17*) and *B. subtilis* (*18*) cells were transformed as described.

Oligonucleotide directed mutagenesis was performed by the method of Kunkel (*19*). Template DNA used for mutagenesis was a 539 bp *Kpn*I-*Pst*I fragment of the *B. stearothermophilus* α-amylase gene cloned into the corresponding sites of M13mp18 or mp19. The oligonucleotides were hybridized with the uridinylated templates extracted from *E. coli* RZ1032. Plaque lift hybridization was used to screen for the mutants of *E. coli* JM103 and the positive candidates were analyzed by sequencing. To reconstruct the α-amylase gene the mutated *Kpn*I-*Pst*I fragment was

[1] This work was supported by Neste Oy Foundation and the Finnish Technology Development Centre.

[2] To whom correspondence should be addressed: M. Vihinen, Department of Biochemistry, University of Turku, SF-20500 Turku, Finland; Tel. +358-21-6335758; Telefax: +358-21-331167; Earn/Bitnet Address: MVIHINEN@FINABO.

Abbreviations: BAA, *Bacillus amyloliquefaciens* α-amylase; BLA, *Bacillus licheniformis* α-amylase; BStA, *Bacillus stearothermophilus* α-amylase; PPA, porcine pancreatic α-amylase; TAA, Taka-amylase A of *Aspergillus oryzae*.

subcloned in pCSS6 and transformed in *E. coli* JM103. The mutated α-amylase genes in pCSS6 were cut with *Dra*I, and the ends were filled in. Then the genes were digested with *Hind*III for cloning in *B. subtilis* with p602/15HT, which had been cleaved by *Bam*HI and *Hind*III after the ends had been filled in. To clone the mutant Arg-232-Lys the wild-type α-amylase gene had to be cloned in p602/15HT and the mutated *Kpn*I-*Pst*I fragment ligated to this construction because the mutation generated a new recognition site for *Hind*III. After ligation *E. coli* MC1061 cells were transformed and the extracted shuttle plasmids were introduced into *B. subtilis* 1A40.

Purification and Characterization of α-Amylase—The wild-type and mutant enzymes were purified from *B. subtilis* clones. The culture fluid was heat treated (75°C, 1 h), after which the enzyme in supernatant was absorbed to insoluble starch (1-5%) in 10% $(NH_4)_2SO_4$. Impurities were removed by washing with 10 mM Tris (pH 7.5), 10% $(NH_4)_2SO_4$; the enzyme was liberated with 4 M urea, 10 mM Tris (pH 7.5) and dialyzed against 10 mM Tris (pH 7.5). 10 mM $CaCl_2$ was included throughout the purification procedure. The incubation time in temperature and pH optimum measurements was 15 min. For the pH optimum determination aliquots of enzymes were incubated at 70°C with substrates buffered to various pH values. The pH electrode used was a Radiometer GK2401C (Radiometer A/S, Copenhagen, Denmark). Complete hydrolysis of starch was performed overnight at 50°C in 10 mM Tris (pH 6.8), 2 mM $CaCl_2$, and 2% corn starch. Remaining starch was removed by centrifugation and samples of the supernatants were analyzed on HPLC in a μBondaback carbohydrate column.

Labeling and Immunoprecipitation Conditions—All the growth media were supplemented with kanamycin. Overnight cultures of *B. subtilis* in 2xTY were used to inoculate SMS medium (*20*). The cells were incubated at 37°C to A_{600} 0.5-0.6 and a 1 ml portion was centrifuged and the cells resuspended in SMS medium lacking methionine. After 30 min [^{35}S]methionine (100 μCi) was added and the cells were allowed to grow for 2 h. The culture was centrifuged to separate cells. The supernatant was saved (extracellular fraction). The cells were washed twice with 10 mM Tris (pH 7.3), 150 mM NaCl, and suspended in 10 mM Tris (pH 7.3), 1 mM EDTA, and then disrupted by sonication. The suspension was centrifuged at 13,000 × *g* for 30 min and the supernatant was saved (cytoplasmic fraction). The immunoprecipitation was performed according to Minsky *et al.* (*21*).

Sequence Comparison and Computer Modeling—Sequence analyses were performed according to Smith and Waterman (*22*) and Needleman and Wunsch (*23*) using GCG software (*24*) and the multiple sequence comparison was according to Vihinen (*25*). The structural models for the deduced amino acid sequences of α-amylases from *B. stearothermophilus* (*6*), *Bacillus licheniformis* (BLA) (*26*), and *Bacillus amyloliquefaciens* (BAA) (*27*) were built by computer modeling on an Evans & Sutherland PS330 graphics terminal. Computer programs employed were Frodo (*28*), Hydra (Polygen, Waltham, Mass.), Insight (Biosym Technologies, San Diego, Calif.), and Gromos (Biomos b.v., Groningen, The Netherlands). The three-dimensional structure of TAA (*4*) was obtained from the Brookhaven Data Bank. Replacements, insertions and deletions were introduced to the TAA structure and the models were refined by energy minimization (*29*).

RESULTS

Sequence Analysis—Numerous α-amylase sequences from widely differing genera have been reported and the refined structure of the TAA at 3 Å resolution is available. Several methods have been employed to compare and align sequences. Recently programs to handle simultaneously several sequences have become available, and the comparisons have been extended to predicted secondary structures (*25*). Comparison and alignment of the amino acid sequence of BStA with that of plant, mammalian, and microbial α-amylases were performed with several computer programs. The results compiled in Fig. 1 show the existence of the four conserved regions in TAA, BStA, BAA, and BLA sequences as reported previously (*2*, *3*). The overall alignment of the sequences is different from that presented by Nakajima *et al.* (*30*).

	I		II	
TAA	117 DVVANH	122	202 GLRIDTVKH	210
BStA	101 DVVFDH	106	230 GFRIDAVKH	238
BAA	98 DVVLNH	103	227 GFRIDAAKH	235
BLA	100 DVVINH	105	227 GFRLDAVKH	235

	III		IV	
TAA	229 GEVLD	233	292 FVENHD	297
BStA	263 GEYWS	267	326 FVDNHD	331
BAA	260 AEYWQ	264	323 FVENHD	328
BLA	260 AEYWQ	264	323 FVDNHD	328

Fig. 1. The four conserved regions of *A. oryzae* (TAA), *B. stearothermophilus* (BStA), *B. amyloliquefaciens* (BAA), and *B. licheniformis* (BLA) α-amylases.

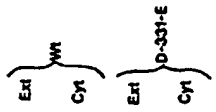

Fig. 2. Synthesis of native (wt) and mutant Asp-331-Glu α-amylases in *B. subtilis*. Cells were labeled with [^{35}S]methionine, centrifuged, and sonicated. The culture fluid (Ext) and disrupted cells (Cyt) were analyzed by 10-15% SDS gel electrophoresis. The enzyme and the fraction are listed above each lane. The arrow indicates position of α-amylase.

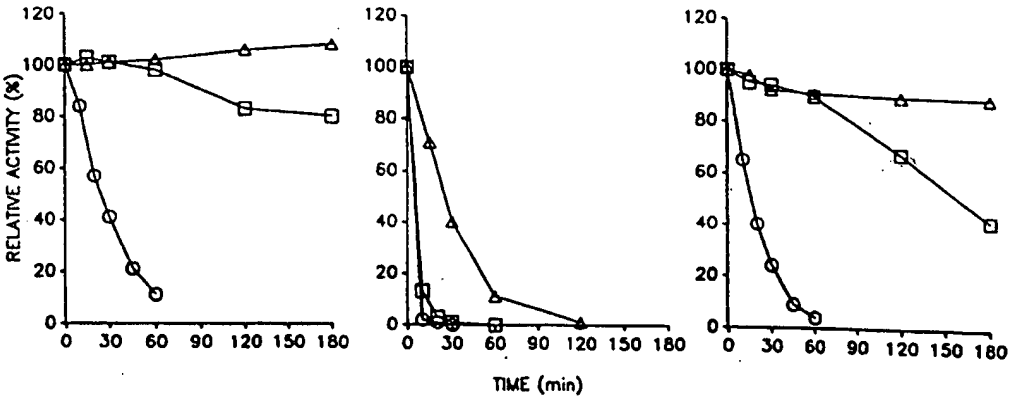

Fig. 3. Thermostability of native (on the left), mutant His-238-Asp (in the middle), and mutant Arg-232-Lys (on the right) enzymes produced in *B. subtilis*. Aliquots of enzymes were incubated at 70°C (△), 80°C (□), and 90°C (○), and samples were withdrawn at the indicated times for residual activity analysis.

TABLE I. Short chain end-products liberated from soluble starch by the native and mutant α-amylases. The end-products are denoted by G1 to G6 to show linear oligosaccharides from glucose to maltohexaose. Total activity indicates the total amount of short chain end-products liberated by μg of protein compared to the wild-type enzyme. The incubation was overnight at 50°C.

Enzyme	Percentage of end products						Total activity
	G1	G2	G3	G4	G5	G6	
Wild type	5.7	21.9	26.7	10.3	21.9		100.0
Arg-232-Lys	1.9	14.5	19.0	6.8	31.1	18.6	58.0
His-238-Asp	22.4	21.5	10.5	5.9	24.3		99.3
Asp-331-Glu		46.5	53.5				15.5

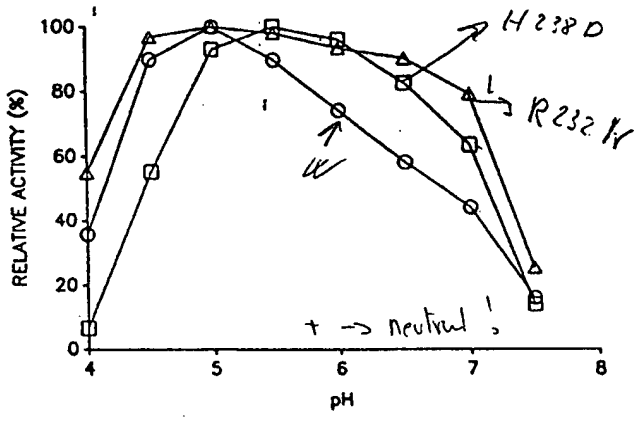

Fig. 4. Activity of the native (○), mutant His-238-Asp (□), and mutant Arg-232-Lys (△) relative to pH at 70°C.

Characterization of Mutant Proteins—The mutagenesis strategy used was that described by Kunkel (*19*). The following mutations were constructed to test the functions of the three most conserved amino acids that were identified by sequence comparison: Arg-232-Lys, His-238-Asp and Asp-331-Glu. The mutant proteins were purified from *B. subtilis* and their enzymatic properties were compared. Immunoprecipitation of the [^{35}S]methionine labeled cell fractions and culture fluid by anti-α-amylase indicated that *B. subtilis* cells containing the wild-type and the mutant Asp-331-Glu α-amylases contained approximately the same amount of precipitating material (Fig. 2). It can be suggested that these proteins are produced at similar rate because the constructions were similar.

Enzymatic Properties of Mutants—Temperature optima of the wild-type and mutant proteins were found to be identical, at 80°C (data not shown). The mutant Arg-232-Lys retained almost the same heat stability as the wild-type α-amylase whereas the mutant His-238-Asp was inactivated remarkably faster (Fig. 3). The pH optima of the parental and mutated proteins are at pH 5.5 (Fig. 4). However, the mutant Arg-232-Lys has a much broader optimal pH for activity, from pH 4.5 to 7.0, than the wild-type enzyme, which has optimal activity from pH 4.5 to 5.5. The pH optimum of the mutant His-238-Asp shifted slightly towards neutrality.

The mutant Arg-232-Lys had only 12% and the mutant His-238-Asp 42% of the specific activity of the parental enzyme. The mutant Asp-331-Glu has almost no activity. The effect of the mutations on the extent of the hydrolysis reaction was tested by analyzing liberated short-chain oligosaccharides. The end-product profile of the mutant His-238-Asp was clearly different from that of the wild-type enzyme and shifted towards shorter end-products (Table I). The mutant enzyme liberated four times more glucose than the parental enzyme. Also the mutant Arg-232-Lys had a slightly altered end-product profile. It gave a higher proportion of longer oligosaccharides, maltopentaose, and maltohexaose. End-product analysis showed that the mutant Asp-331-Glu also exhibited some amylolytic activity, the only short-chain products being maltose and maltotriose. The yield of oligosaccharides was only 15.5% of that of the wild-type enzyme. The mutant His-238-Asp liberated similar amount of short-chain saccharides as the wild-type enzyme and the mutant Arg-232-Lys about half of that. The specific activities and total activities in Table I cannot be compared, since the total activity is given at end point.

The effect of soluble starch concentration on the activity was examined at 70°C and pH 5.5. The Michaelis constant is 4.5, 5.3, and 6.2 mg/ml for the wild type, Arg-232-Lys and His-238-Asp mutants, respectively.

Modeling of α-Amylase Structures—The three-dimensional structures of BStA, BAA, and BLA were modeled based on the TAA structure in an attempt to understand the

relation between the altered properties of the mutant proteins and their structure. The entire structure of the *Bacillus* α-amylases could not be modeled from TAA because they all have an insertion of about 45 amino acids near the active center and the Ca^{2+} binding site. The insertion is unique to liquefying *Bacillus* α-amylases; it cannot be found in any other α-amylase sequence published. This region could perhaps participate in substrate binding. The *B. amyloliquefaciens* and the *B. licheniformis* α-amylases were modeled to aid in evaluating the validity of the *B. stearothermophilus* model, since homology between BStA and TAA is not more than about 40%. The *Bacillus* α-amylases are more similar: homology is 60-70%. If independent models of quite homologous α-amylases are essentially similar, the validity of the models is supported. The active site regions are very similar in all three models (*29*). The overall structure of PPA is similar to that of TAA. Unfortunately only the coordinates of TAA were available.

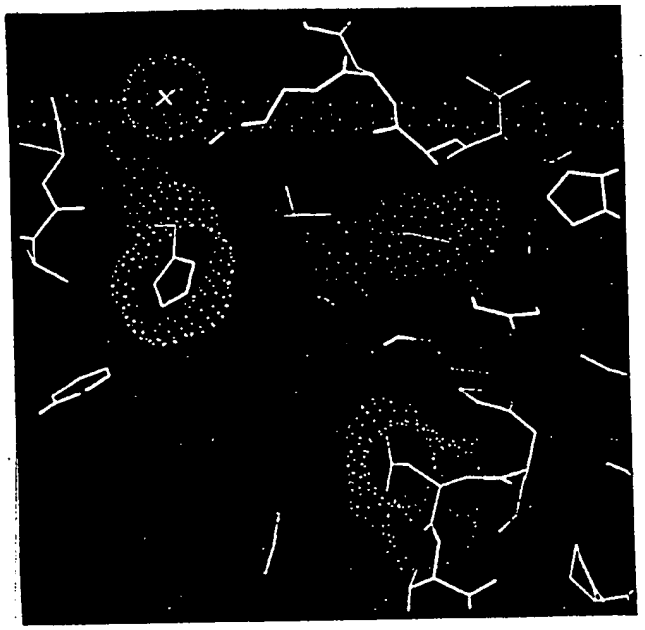

Fig. 5. **Computer model of BStA catalytic residues and His-238.** The model shows the function of His-238 (in yellow) in substrate binding and Ca^{2+} binding. Ca^{2+} is shown in the upper left corner with magenta dots. The surface of the active site, Asp-234 (upper) and Asp-331 (lower), is indicated by blue dots. Maltotriose as substrate is in red. The glucosidic bond to be broken is in white.

DISCUSSION

The importance of three conserved residues in thermostable *B. stearothermophilus* α-amylase was studied by site directed mutagenesis. Computer methods were used to explain the results. Even when the protein structure has not been determined, as was the case in this work, computer modeling can be useful and suggest hypotheses which can further be tested by mutagenesis studies.

One of the four conserved regions in α-amylases, region I, has been suggested to be involved in Ca^{2+} binding (*2*). It has been proposed that Asn-121 of this region in TAA is a Ca^{2+}-binding residue (*5*). The other amino acids related to Ca^{2+} binding in TAA are Asp-163, Asp-175, and His-210 (*5*). The His-210 is also involved in substrate binding and it is conserved in region II, but the two other amino acids, although quite conserved, are not located in the conserved regions. Based on the structural model, the Ca^{2+} is bound to BStA by side chain atoms of Asp-105, Asp-190, and Asp-203 and by main chain atoms (probably *via* carbonyl oxygen) of His-238 (*29*). The structure of the TAA and the

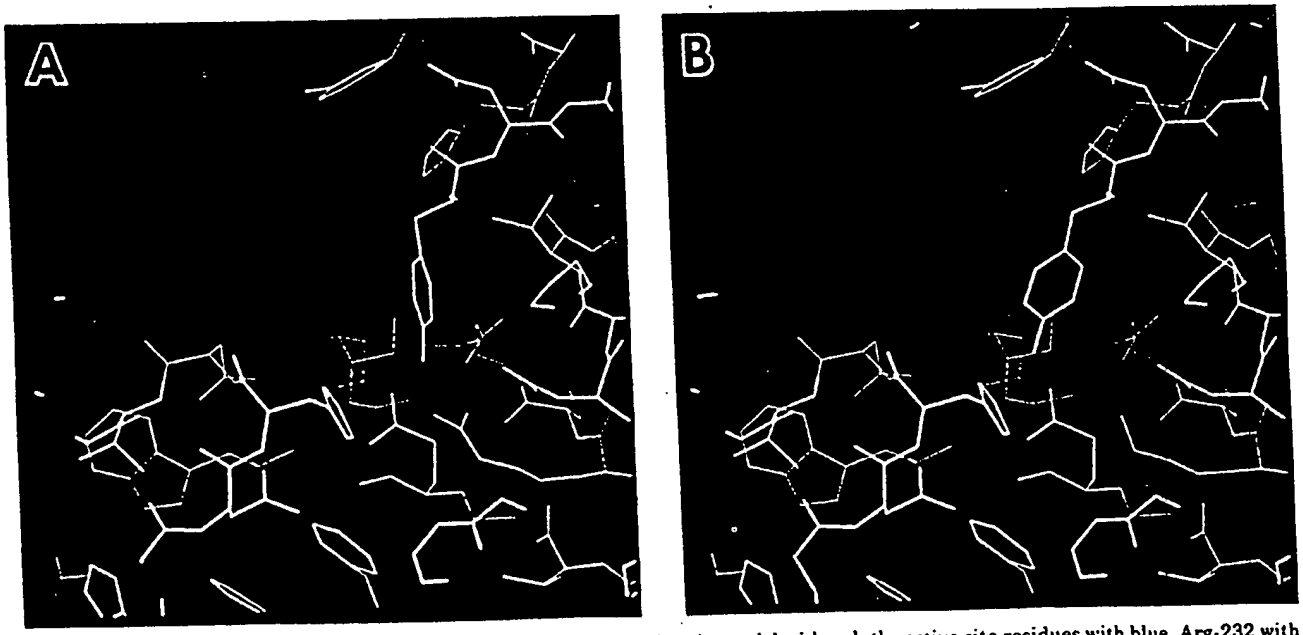

Fig. 6. **The proposed action of Arg-232.** The substrate is indicated in the model with red, the active site residues with blue, Arg-232 with magenta, and Tyr-63 with yellow. A: Native Arg-232 in interaction with Tyr-63. B: The mutant Arg-232 Lys cannot form a salt bridge with Tyr-63, which is free to prevent binding of the substrate to the active site.

models support the view that region I is necessary for maintaining the conformation of domain B (nomenclature according to Ref. 5) suitable for forming the Ca^{2+} binding site. Many of the conserved residues from regions I and II are involved in forming the interface between domains A and B. The function of the Ca^{2+} could be in tightening the connection of the domains. Regions II and III are involved in forming the active site and substrate binding regions. Based on the sequence comparisons and the modeled structures, we propose that Asp-234 and Asp-331 are the active site residues in BStA.

Mutant Asp-331-Glu—According to our model, the substrate is bound to BStA between the Asp-234 and Asp-331 (Fig. 5). The distance between the bound substrate and the carboxyl oxygen atoms of the active site aspartates is in native enzyme 3 Å, but in the mutant Asp-331-Glu it is only 2.2 Å. The shorter distance between the active site residues in the mutant protein would appear to cause a steric hindrance and does not leave enough space for the substrate to bind correctly.

This mutant protein was not analyzed because of much reduced activity. Maltose and maltotriose are the only short-chain end-products liberated from starch by this mutant, possibly because of a changed substrate binding mechanism. The reduced activity of the mutant protein is probably not due to low stability because it is not more active in lower temperatures (data not shown). Immunoprecipitation of mutant protein indicates that the gene for the mutant Asp-331-Glu is expressed as efficiently as the parental enzyme. From Fig. 2 it can be seen that the enzyme produced is efficiently secreted to the extracellular medium, which rules out the possibility of restricted export.

Mutant His-238-Asp—His-210 in TAA has been suggested to be involved both in substrate binding (*4*) and calcium binding (*5*). The sequence comparison shows that the counterpart in BStA is His-238, and the model suggests that it could indeed participate in both functions, too. It was replaced by Asp to introduce a charge change near the active site.

Figure 5 shows the substrate binding to His-238 and catalytic site residues. The substrate is bound to the active site cleft by hydrogen bonds and hydrophobic interactions (*5*). The His-238 could be one of the key interaction sites outside the catalytic center and may govern the proper orientation of the flexible substrate. The reduced specific activity of the mutant His-238-Asp (42% of that in the parental enzyme), higher amounts of starch required in purification to obtain binding, and shift towards shorter end-products (Table I) suggests that this mutant has an altered substrate binding to the subsites. However, the K_m value of the mutant protein is essentially similar to that of the wild-type enzyme. This may indicate that in the mutant protein binding to subsites outside the active site could be rate limiting although the K_m has changed only slightly. The mutation His-238-Asp has reduced thermostability; activity is almost totally lost at 80°C in 20 min whereas the parental protein retains 80% of the activity even after 2 h under the same conditions (Fig. 3). High Ca^{2+} or Na^+ concentration does not affect the thermal stability. The temperature optimum, however, is the same as the parental enzyme. The mechanism of stabilization by Ca^{2+}, which is bound to the interface of two domains, A and B, is presumably by tightening the connection between them. His-238 seems to interact with the cation with the main chain carbonyl oxygen as in TAA and PPA (*5*). Because almost any residue could provide this ligand for the Ca^{2+} binding, the side chain must also be important for the function of this residue.

The negative charge of His-238-Asp causes electrostatic repulsion which could be compensated at least to a certain extent by change in the dihedral angles of the backbone and the side chain atoms. The lowered thermal stability of the mutant may be due to the reduced possibility of the main chain carbonyl oxygen contributing to Ca^{2+} binding. This leads to a more open flexible structure which could cause reduced thermal stability (*31*) by inducing a more unfolded structure (*32, 33*).

Mutant Arg-232-Lys—Sequence comparisons have identified highly conserved Arg-232, for which no function has been suggested. It was therefore interesting to test the function of this amino acid by directed mutagenesis. The conservative mutation Arg to Lys was made, having the same length but missing the strongly basic δ-guanido group.

The mutant Arg-232-Lys has about the same pH and temperature optimum as the parental protein but a slightly lower thermostability (Figs. 3 and 4). The end-product profile is essentially similar to the parental enzyme (Table I). The most important difference is lower specific activity, only 12% of that found in the wild-type enzyme. This result is important because this amino acid was not previously thought to affect activity or substrate binding. The structural model suggests that it is near but not bound to the substrate. However, it must contribute to function because the conservative mutation remarkably reduced the specific activity.

We suggest that the Arg-232 is needed to withdraw Tyr-63 from the active site by ionic interaction. Sequence comparison shows that Tyr-63 and Arg-232 are conserved amino acids. The Tyr residue protrudes in a similar fashion into the active site region of TAA as in the structural models of the three *Bacillus* amylases. This may be due to the energy minimization used in the refinement of the TAA structure (*4*), which has turned the side chain towards the active site cleft. The proposed substrate binding model of TAA does not suggest that Tyr-82 is involved in substrate binding in TAA. However, it does not rule out the possibility that Tyr-63 forms a part of the "wall" of the active site cleft in BStA. We suggest that this residue interacts electrostatically with Arg-232 (Fig. 6A) so that the basic Arg binds to the hydroxyl oxygen of the phenyl group and orientates the side chain so that the active center is free to bind the substrate. The side chain of Arg is much more basic than that of Lys; thus the binding in the mutant is weaker and the Tyr-63 is free to rotate and so disturb the active center, reducing the specific activity (Fig. 6B). Also here the K_m value is essentially similar to that for the parental enzyme. When Tyr-63 is bound to Lys-232, the substrate binding and K_m should be similar to the wild-type enzyme, but when unbound the substrate binding is hindered.

It is possible that Arg-232 is also connected to the nearby Asp-117 residue in the conserved region I. It is also possible that the highly basic Arg-232 could interact with both these residues and stabilize the structure of the enzyme by

decreasing the motility and flexibility of domain B. The mutant Arg-232-Lys cannot provide the basic guanido group of Arg for binding the other two side chains. Thus, the somewhat lowered thermal stability is probably due to less effective binding, which permits increased flexibility.

This work supports the idea that the identification and study of conserved amino acids by sequence comparison, even in the absence of known tertiary structure, can lead to useful insights on the function and structure of proteins.

We are grateful for Dr. Tuula Teeri and Dr. Leif Laaksonen for establishing a computer aided molecular modeling facility at VTT.

REFERENCES

1. Vihinen, M. & Mäntsälä, P. (1989) *Crit. Rev. Biochem. Mol. Biol.* 24, 329-418
2. Rogers, J.C. (1985) *Biochem. Biophys. Res. Commun.* 128, 470-476
3. Mackay, R.M., Baird, S., Dove, M.J., Erratt, J.A., Gines, M., Moranelli, F., Nasim, A., Willick, G.E., Yaguchi, M., & Seligy, V.L. (1985) *BioSystems* 18, 279-292
4. Matsuura, Y., Kusunoki, M., Harada, W., & Kakudo, M. (1984) *J. Biochem.* 95, 697-702
5. Buisson, G., Duee, E., Haser, R., & Payan, F. (1987) *EMBO J.* 6, 3909-3916
6. Suominen, I., Karp, M., Lautamo, J., Knowles, J., & Mäntsälä, P. (1987) in *Extracellular Enzymes of Microorganisms* (Chaloupka, J. & Krumphanzl, J., eds.) pp. 129-137, Plenum Press, New York
7. Suominen, I., Karp, M., Lähde, M., Kopio, A., Glumoff, T., Meyer, P., & Mäntsälä, P. (1987) *Gene* 61, 165-176
8. Suominen, I., Käpylä, J., Tilgmann, C., Glumoff, V., & Mäntsälä, P. (1988) *FEMS Microbiol. Lett.* 55, 3-8
9. Reinikainen, P., Lähde, M., Karp, M., Suominen, I., Markkanen, P., & Mäntsälä, P. (1988) *Biotech. Lett.* 10, 149-154
10. Vieira, J. & Messing, J. (1982) *Gene* 19, 259-268
11. Casadaban, M.J. & Cohen, S.N. (1980) *J. Mol. Biol.* 138, 179-207
12. Kunkel, T.A., Roberts, J.D., & Zakour, R.A. (1987) *Methods Enzymol.* 154, 367-382
13. Yanisch-Perron, C., Vieira, J., & Messing, J. (1985) *Gene* 33, 103-119
14. Maniatis, T., Fritsch, E.F., & Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. Cookson, C. (1978) *Anal. Biochem.* 88, 340-343
16. Zagursky, R.J., Berman, M.L., Baumeister, K., & Lomax, N. (1985) *Gene Anal. Tech.* 2, 89-94
17. Norgaard, M.V., Keem, K., & Monahan, J.J. (1978) *Gene* 3, 279-292
18. Gryczan, T.J., Contente, S., & Dubnau, D. (1978) *J. Bacteriol.* 134, 318-329
19. Kunkel, T.A. (1985) *Proc. Natl. Acad. Sci. U.S.* 82, 488-492
20. Anagnostopoulos, C. & Spizizen, J. (1961) *J. Bacteriol.* 81, 741-746
21. Minsky, A., Summers, R.G., & Knowles, J.R. (1986) *Proc. Natl. Acad. Sci. U.S.* 83, 4180-4184
22. Smith, T.F. & Waterman, M.S. (1981) *Adv. Appl. Math.* 2, 482-489
23. Needleman, S.B. & Wunsch, C.D. (1970) *J. Mol. Biol.* 48, 443-453
24. Devereux, J., Haeberli, P., & Smithies, O. (1984) *Nucleic Acids Res.* 12, 387-395
25. Vihinen, M. (1988) *Comput. Appl. Biosci.* 4, 89-92
26. Yuuki, T., Nomura, T., Tezuka, H., Tsuboi, A., Yamagata, H., Tsukagoshi, N., & Udaka, S. (1985) *J. Biochem.* 98, 1147-1156
27. Takkinen, K., Pettersson, R.F., Kalkkinen, N., Palva, I., Söderlund, H., & Käärikäinen, L. (1983) *J. Biol. Chem.* 258, 1007-1013
28. Jones, T.A. (1978) *J. Appl. Cryst.* 11, 268-272
29. Vihinen, M. & Mäntsälä, P. *Biochem. Biophys. Res. Commun.* in press
30. Nakajima, R., Imanaka, T., & Aiba, S. (1986) *Appl. Microbiol. Biotech.* 23, 355-360
31. Vihinen, M. (1987) *Prot. Eng.* 1, 477-480
32. Tomazic, S.J. & Klibanov, A. (1988) *J. Biol. Chem.* 263, 3086-3091
33. Tomazic, S.J. & Klibanov, A. (1988) *J. Biol. Chem.* 263, 3092-3096